Ph.D. Final Synopsis

Title: "Peopling of Kavaratti Island of Lakshadweep – A Population Genetic Approach"

Reg: MU/EXB/Ph.D./CR.26/Biotech./2021-22/E.25

Research Scholar:
Alnoman Mundher Tayyeh
Department of Biosciences
Mangalore University
Mangalagangothri

Research Supervisor
Dr. Mohammed Shafiul Mustak
Professor
Department of Applied Zoology
Mangalore University
Mangalagangothri

Place of Investigation:

Department of Biosciences, Mangalore University,
Mangalagangothri, Mangaluru- 574199

October - 2023

Ph.D. Final Synopsis

Synopsis of the thesis submitted for the award of Ph.D. degree in Biotechnology of Mangalore University, Mangalagangothri, Karnataka, India.

Title of thesis-

"Peopling of Kavaratti Island of Lakshadweep - A Population Genetic Approach"

Candidate: Alnoman Mundher Tayyeh

Introduction

The genome sequences and population differences have provided plenty of opportunity to examine the genes and evolutionary- and medically-important locations. The genesis and spread of distinct populations have been better understood from an evolutionary standpoint thanks to the genes and genomes. Human population Geneticists and archaeologists/anthropologists took advantage of the opportunities to use these contemporary tools to understand how human populations in various continents, nations, and regions of the world have differed from one another as well as how, when, and where our species' origins occurred (Bamshad et al., 2001; Mellars et al., 2013). Thus, many important discoveries and the development of numerous technologies, such as, heralded the start of the genomic era.

It has become crucial to consider the terrain, archaeology, culture, and language of the studied area in order to better comprehend the conceptual framework of the issue at hand and to define the origins, expansion, and movement of populations in the study regions. I briefly review them before moving on to the matter (Thangaraj et al., 2003; Sahoo et al., 2006). India has the reputation of supposedly having the most biological and racial diversity. and due to credit for its biological diversity, even the country's location at the confluence of the African, Oriental realms, and northern Eurasian (Majumder, 2001). This biological wealthiness has attracted many streams of people to the subcontinent at various times and from multiple directions, resulting in the mixing of a wide variety of human genes and cultures. In India, as opposed to other places where dominant human cultures have a tendency to ingest or eradicate others, the trend in India has been to isolate and dominate the subordinated. Measuring traits like social, ethnic, cultural architecture, and linguistic as a result of demographic shifts defines a population structure.

Homo sapiens spread about 300,000 years before, at the period of its speciation. The latest hypothesis of African origin proposes that anatomically modern human beings outside Africa are descended via a group of Homo sapiens originating in East Africa around 70-50,000 years ago and extending around 50,000 years ago across the South Asian coast of Oceania. Contemporary humans have spread all over Europe for 40,000 years, (Schlebusch et al., 2017). Middle and late Miocene hominoids were found not only in Africa but mostly in regions of Eurasia, our supposed ancestors living in the grasslands of South and East Africa, including the relatives of our Homo species and the great apes, Homo erectus appear to be the first Homo discovered peacefully outside Africa approximately 2 million years ago. Human beings have continued to occupy a large amount of space in this world. Nevertheless, 200,000 years ago, *Homo sapiens* was just a newly budding species in Africa, while relatives such as Homo erectus and Homo heidelbergensis already had dispersed to explore areas of Eurasia outside Africa. (Groeneveld, 2017).

It is estimated that anatomically modern humans first appeared in the Indian subcontinent between 60 and 50 kybp (years before present) (Bittles, 2002). India has played a significant role in the movement of modern people, who originated in Africa some 100,000 years ago (Cann, 2001). Remains of contemporary humans from the late Pleistocene period (55000–25000 ybp). Humans appear to have expanded their range in several areas of India by the middle Paleolithic epoch (50,000–20,000 ybp) (Misra, 2001). The event of India's demographic explosion has been pushed to 60,000-85,000 ybp by molecular genetic evidence (Ottolenghi et al., 1995). Perhaps the population of Southeast Asia and, later (50,000–60,000 ybp), Australia was affected by this increase and subsequent migration. Another independent migration of modern humans is thought to have occurred in southern China around 60,000 ybp, and it may have contributed to human migration into Southeast Asia and India (Ballinger et al., 1992).

Kavaratti island is one of the oldest ten inhabitant islands of Lakshadweep. Kavaratti Island is the capital of Lakshadweep which is under the Union Territory of India. Kavaratti is located in the Arabian Sea, approximately 215 miles (345 kilometres) west-southwest of Calicut city and also the Malabar Coast of southern India. On the western side of the island. According to the latest population estimates in 2011, Kavaratti had a total population of 11210, (Ranith et al., 2017). In the current study attempted to use the Y chromosome and mtDNA as a tool to investigate the ancestral origin and migratory route of this isolated Kavaratti Island.

Review of Literature

The most recent study from David Reich lab showed that in the last 10,000 years, there had been two big migrations into India. The first migration came from south-western Iran's Zagros region date back to 7,000 and 3,000 BCE, and took farmers, most likely herders, to India. Later these Zagrosian farmers mixed with the subcontinent's earlier inhabitants the first Indians, descendants of refugees from Out of Africa (OoA) who had arrived in India 65,000 years ago and together they formed the Harappan civilization. The second group of immigrants (the Aryans) from the Eurasian Steppe came in the centuries after 2000 BCE, maybe from the region now known as Kazakhstan. They probably took across an early version of Sanskrit, horse knowledge and a variety of new cultural traditions, which all laid the foundation for early Hindu/Vedic society, including such sacrificial rituals. A thousand years earlier people had also moved from the Steppe into Europe, to replace and mix with farmers there, spawning new cultures and spreading Indo-European languages (Joseph, 2018; Narasimhan et al. 2019).

Haplogroups are used to represent the major branch points on the mitochondrial phylogenetic tree. Understanding the evolutionary path of the female lineage has helped population geneticists trace the matrilineal inheritance of modern humans back to human origins in Africa and the subsequent spread around the globe. Individual mtDNA macro haplogroups are widely circulated across various regions, each with a distinct pattern in contrast to the others. India, in especially, has the most varied mtDNA Hg M, N, and L distribution pattern in SA (Basu et al., 2003; Metspalu et al., 2004; Palanichamy et al., 2004; Quintana et al., 2004; Thanseem et al., 2006; Chaubey et al., 2007). This greater diversity could be attributed to the subcontinent playing a significant role in the colonization and dispersal of humans

India, which represents incredibly diversified deep root lineages and displays region-specific distinctiveness among the Indian population, has the highest prevalence of the Macro Hg M (Mustak et al., 2019). India is considered to have the second-most diverse population in the world after Africa owing to the greater diversity of its mtDNA, which more accurately represents the region's unique demographics (Pasupuleti et al., 2020). According to Palinchami et al., 2004), Hg R is the second-most common mtDNA Hg behind M in the Indian subcontinent and it is inherited from the West Eurasia (WE) Hg macrohaplogroup N. (Kivisild et al., 1999a).

There were no reports that could identify the mtDNA lineages nested in the Hg R until 2004 (Metspalu et al., 2004). The entire mtDNA sequences including new sub Hgs of macro-Hg R (R5, R6, R7, R8, R30 and R31) were later disclosed by Palanichamy et al., 2003.

There are no studies till date on the ancestral origin and migration route of the Kavaratti Islands, therefore, the following objectives were framed to determine the tracing of the Kavaratti Island population.

Objectives:

1. To understand the maternal genetic ancestry of the Kavaratti Island population of Lakshadweep.
2. To understand the paternal genetic ancestry of the Kavaratti Island population of Lakshadweep.
3. To find the Genetic affinity between Kavaratti Island and other islands of Lakshadweep.
4. To find the Genetic affinity of Kavaratti Island population with neighbouring South Asian population.

Materials and Methods

Sample:

About 5mL-9mL of blood samples were randomly collected from 150 unrelated males residing on Kavaratti Island. Informed consent was obtained from the participants. The study was conducted after obtaining the approval from the Institutional Human Ethics Committee, Mangalore University as per the IHEC guidelines (MU/AZ/20/IHEC/2014-2015 dated 07/04/2014).

DNA extraction and genotyping:

DNA was extracted using the standard Phenol Chloroform method. The partial coding region (14856- 16023), HVR-I (1-576), and HVR-II (16024-16569) of mtDNA were amplified by an ABI Veriety 96-well (Applied BioSystems) Thermocycler. The mtDNA region was amplified and genotyped using PCR. The amplified PCR products were sequenced using Sanger's Di-deoxy chain terminator cycle sequencing method.

mtDNA sequences were aligned against the reverse Cambridge Reference Sequence (rCRS) in Ugene Unipro Software (Okonechnikov *et al.*, 2012). Variations were scored and haplogroups were assigned according to Phylotree 17 FU1 (Dür, Huber and Parson, 2021), using Haplogroup 2.4.0 software (Weissensteiner *et al.*, 2016), followed by a EMPOP database query check (Huber, Parson and Dür, 2018). Two samples showing ambiguous results were excluded from the analysis.

Statistical analysis:

DnaSP6 software (Rozas *et al.*, 2017) was used to calculate Nucleotide Diversity, Haplotype Diversity and Tajima's D indices, based on sequence variation. For comparative analysis, pairwise Fst among different populations was calculated using Arlequin Software (Excoffier and Lischer, 2010). The same software was used for AMOVA (Analysis of Molecular Variance). PAST4 software (Hammer, Harper and Ryan, 2001) was used for multivariate analysis, such as Principal Component Analysis (PCA) and Neighbour Joining Tree clustering. M2 and R30 haplogroup network analysis was performed using the POPART package in R (Leigh and Bryant, 2015). The Phylogeny Tree was constructed using Mega 11 software (Tamura, Stecher and Kumar, 2021).

Bayesian analysis:

Bayesian age estimation was performed using BEAST v2.7.4. Samples were filtered based on the haplogroups. Ancient mitogenomes (with higher coverage and C14 dates) from AmtDB were used for calibration (Ehler et al. 2019). The L2c2 mitogenome from a Moreno individual (PaMOR16007) was used as an outgroup. BEAST runs were performed using the non-coding region sequences. Trees were linked, and runs were performed with the HKY substitution model. A strict molecular clock was chosen with mutation rates of 1.292 and 0.369 mutation/site/million years for the HVS1 and HVS2 regions respectively (Connell et al. 2022).

A consensus tree was generated using Tree annotator with the following settings: 50% burnin, Common Ancestor Heights. The consensus tree was prunned in iTOL. In Figtree, the tree was rerooted to L2c2 outgroup.

Median Heights with 95% HPD (High Probability Densities) were considered as TMRCAs and reported. All the runs were performed with 5,000,000 burnin and 50,000,000 iterations, drawing samples at intervals of 10,000 MCMC (Markov chain Monte Carlo) steps (Olivieri et al. 2017; Brandini et al. 2018; Capodiferro et al. 2021).

Results and discussion

Mitochondrial Genomic diversity in Kavaratti islanders:

Haplogroups R30b2 (32.1%), M2a'b (17.7%), M33a (10.3%), M3a1+204 (9.0%) and U4b1a1a1 (5.1%) account for about 70% of the total mtDNA variation on Kavaratti Island. The abundance of macrohaplogroup M (51%) in this population resembles mainland populations of the Indian subcontinent (Anthropological Survey of India, 2021). Macrohaplogroup R accounted for 32% of the total distribution. The U4 haplogroup occurred at 10% frequency. Viewed at the macro level, this population carries Indian-specific M and ancient N in relatively equal proportion. These two haplogroups contrast with the mainland tribal gene pool, which shows a significant difference between the two. Interestingly, such equal distribution is also a characteristic of Northern India (Anthropological Survey of India, 2021).

Furthermore, Bayesian analysis of M2 sequences from the available databases clustered the M2 subclade found in Kavaratti Island separately. This Kavaratti M2 cluster shares a common ancestor with the northwestern M2 clade and appears to be younger (41,611 YBP, 95% HPD: 18,983 – 68,032) than the other clades found in the Indian subcontinent with coalescence times of 46 to 57 thousand years. Earlier studies have reported a broader range for the time to the most recent common ancestor for M2 subclades in India, with M2a dated between 21 and 50 thousand years and M2b dated between 12 and 55 thousand years. Nevertheless, with the exception of M2, which seems to be absent in our samples, the time to the most recent common ancestor for M2a fell in the pre-Last Glacial Maximum period. This time frame for the Kavaratti M2 cluster in addition to its affinity towards the northwestern M2 clade suggests that our samples share M2 ancestry with the northwestern populations.

Diversity statistics:

Nucleotide diversity (π) in the Kavaratti population is 0.00122, while the haplotype diversity (Hd) is 0.993. A lower π index and a higher Hd index suggest a population bottleneck and a rapid population expansion with accumulation of mutations. A statistically significant Tajima's D value -1.92469 indicates a recent population expansion.

Paternal population structure of Kavaratti Island population

To investigate the origin of paternal lineages of Kavaratti populations, 78 samples, were analysed for 17 alleles markers. A total of 23 unique haplotypes were observed. Our data show great diversity among the population of Kavaratti Island. A large diversity may be due to the fact that Kavaratti Island being the capital of the Lakshadweep archipelago, and a lot of demand for settlement there for various purposes, including trade. The hg R was found to be the most prominent macro haplogroup of the Kavaratti population with a higher frequency of 23 %, followed by haplogroup L was observed as the second most frequent haplogroup in the Kavaratti Island population, which characterizes (19%) of the total Y chromosome gene pool.

Apart from the major two haplogroups observed in Kavaratti Island, there are a few more minor haplogroups observed, the minor haplogroup J accounts for (18%), Followed by O (12%) and E (7%). and the remaining haplogroups detected were less than (6%). The sub-lineages of the Y chromosome macro haplogroup R, such as the haplogroup R1 subspecies including (R1a, R1b), constituted approximately 11.52 % of the total samples, as well as the haplogroup R2 accounted for approximately 11.52% of all samples. Further three subclades of the haplogroup L were also been observed, frequencies are as follows: L1b (11.54%), L1a (6.4%), and L1c (1.2%). In Hg J, four sub-lineages were detected: J2a (11.54%), J2b (2%), J1a (2%) and J1a3 (1%) with lower frequency levels. two sub-lineages belonging to haplogroup O, O1b1 (11%) and O1b2 (1%), were also found. As for haplogroup E, it was found that it represents (7%) of the total samples, and it contained 3 sub-lineages as follows: E1b1b (5%), E1b1a (1%) and the same frequency (1%) for E2. As for the other remaining haplogroups, their lineage was present in a small percentage among the Kavaratti population, respectively: Haplogroup Q constituted 6.4%, G2 5%, H1a1a 3.85%, D 2% and finally I2a2a 1.2% as the lowest haplogroup among Kavaratti population.

The Genetic Affinity between Kavaratti Island and other islands of Lakshadweep

The Maternal Genetic Affinity between Kavaratti Island and other islands of Lakshadweep

The Lakshadweep archipelago population is distributed over 10 islands. To get an overview of Lakshadweep's overall gene pool, we compared their average pairwise differences (Fst) based on haplotype frequencies. The Kavaratti population (Fst = 0.858) is a more diverse group compared to Chetlat (0.324), Bitra (0.564) and Kadmat (0.526).

This enhanced diversity is primarily due to the stronger founder effect of R30 on these islands. The R30 frequency on these islands is >60%, while other more diverse islands including Kavaratti show <40%. The Lakshadweep islands are different form each other in area, vegetation and population. To assess the relationship between founder haplogroups and island area, we performed a Pearson correlation test. Haplogroup R30 is significantly ($p<0.05$) negatively correlated with Fst ($r2 = -0.80$) and island area ($r2 = -0.71$). The result supports the hypothesis that the earliest settlers had a high frequency of R30 when they reached the Lakshadweep archipelago. Similarly, smaller islands have accumulated this haplogroup in high frequency. Haplogroup M2 shows insignificant positive correlation with Fst ($r2 = 0.13$) and island area ($r2 = 0.20$). There is insignificant positive correlation between the island area and Fst ($r2 = 0.54$, $p = 0.1$), i.e., the greater the area, the more diverse the population.

The Paternal Genetic Affinity between Kavaratti Island and other islands of Lakshadweep

We have compared our Y-STR data to better understand the founding paternal lineages present on Kavaratti Island with other islands of Lakshadweep. Kavaratti Island showed a wide diversity of haplogroups, where new parental sub-lineages were observed that are not present or may be minimal on other islands of Lakshadweep. The reason behind this diversity may be attributed to the fact that Kavaratti Island is the capital of the Lakshadweep region and an important commercial centre for people's migration and settlement there. The Lakshadweep islands are different from each other in area, vegetation and population. PCA shows a distinct cluster for Kavaratti. All the northern islands clustered together. It shows that Kavaratti clustered with islands of larger areas. There is an insignificant positive correlation between the island area and diversity, (i.e., the greater the area, the more diverse the population). Interestingly, Kavaratti belonged to the cluster including Minicoy, Androth and Bitra suggesting admixture between the individuals of these islands.

This enhanced diversity is primarily due to the stronger founder effect of Y-chromosome haplogroup R on these islands. NJ Tree of Y-STR haplogroup showed Kavaratti clustered together with Minicoy, Androth and Bitra. By these results, we can summarise that genetic admixture between Kavaratti and these islands occurred more frequently than between Kavaratti and other, far further islands. T

his resemblance may be due to the tight proximity of these islands to Kavaratti. Further, the Kavaratti population is a more diverse group compared to Chetlat, Agatti and Kiltan. The result supports the hypothesis that the earliest settlers had a high frequency of Y-chromosome haplogroups R when they reached the Lakshadweep archipelago.

The Genetic affinity of the Kavaratti Island population with the neighbouring South Asian population

The Maternal Genetic Affinity between Kavaratti Island with the neighbouring South Asian population

To reconstruct the maternal ancestral lineage of the Kavaratti islanders, we compared their gene pool with that of the mainland populations. We used Principal Component Analysis (PCA) to visualise the clustering pattern among them. Interestingly, Kavaratti clustered along with the Nair, neighboured by Kodava, Madia, Koraga, Kathodi and the Punjab Schedule Castes. This cluster lies exactly in between the Western-Central-Eastern cluster and Northern-Northwestern cluster. A similar clustering pattern is seen in the Neighbour Joining (NJ) Tree plotted based on Hamming's dissimilarity and Andamanese root. This clustering is due to the R30, M2, M3, M30 and L haplogroups present in these populations.

There are no significant affinity obtains between the Kavaratti and Sri Lankan populations. However, a Sri Lankan tribe called Vedda clustered closer to Kavaratti than the general population of the island. We observed a similar pattern when larger datasets (1000 genome) were included in the Phylogeny Tree. also In the Phylogeny Tree, Kavaratti islanders (KAV) clustered with the Indo-European individuals.

The multivariate analysis, therefore, suggests the Kavaratti population belonged to a unique wave of migrants that neither belonged to the earliest Indian specific M2 carriers nor to the West Eurasian descendants. A peculiar feature in the Kavaratti cluster is the presence of M3, M30 and U2. M3 and M30 originated during or after the LGM event. This time depth is important, since the members of the Kavaratti cluster carry M3 and M30, leading us to the assumption that these populations had similar gene pools during the LGM recovery, after which bottleneck events took place. At this juncture, if an R30 bottleneck happened, then it is obvious that the group that migrated to Kavaratti belonged to a bigger group including the ancestors of the Urali Kuruman, the Punjab Scheduled Castes, UP Brahmins, Thogataveera, Reddy, Jat Sikh, Nair etc.

Therefore, we hypothesise that the Kavaratti islanders trace their maternal ancestry to the Northwestern region of the Indian subcontinent. To resolve further, based on the subclade information available for the R30b2 and R30 haplogroup network analysis, we suggest that Kavaratti islanders are maternally related to Mazhabi Sikhs, Balmikis (Punjab Scheduled Castes) (Singh et al., 2016) and Nairs. All other groups carry R30a subclades. Interestingly, Nairs have two important characteristics which are similar to Lakshadweep islanders. Firstly, their Malabar roots as described in oral traditions and documented history, and secondly high status is enjoyed by women (Forbes, 1979). However, if the Kavaratti islanders were recent offshoots of the Nairs, then it would most highly have been mentioned in folklore. At present, no record of any direct Nair lineage to the Lakshadweep islanders, written or oral, is known to exist (Chib, 1985).

The Paternal Genetic Affinity between Kavaratti Island with the neighbouring South Asian population

To determine the relationship of the Kavaratti Population with the neighbouring South Asian population, Y-chromosome Hg frequencies were retrieved from the literature and distinguished based on geography and ethnicity. The final data set for Y-chromosome analysis contained South Indian (SI) populations, North Indian (NI), South Asian and Middle East populations. Population-specific can be demonstrated in the para haplogroup or its derivation where the Y-chromosome Hgs are diffusing in the mainland. This may be attributed to antiquity, demic dispersion, and recent bottleneck incidents of the Y-chromosome expansion through the Neolithic period (Sahoo S. et al., 2005; Arunkumar G. et al., 2015). Studies on the Y-chromosome had drawn data indicating that an ethnic group or geographic region would be a representation of a pre-historical bottleneck event. And this will gain more significance with information concerning the geographic distribution of the compatible Y-chromosome haplogroup frequencies.

NMDS plot shows an interesting clustering pattern, we observe that the Y-STR haplotypes of Kavaratti islanders show similarity with the North-western populations like Sindhi, Pathan and Gujjar. Also, affinities do not observe any cluster with the south Indian populations from Tamil Nadu and Andhra Pradesh. Overall, our results suggest a different paternal lineage among the Kavaratti islanders, indicating that the paternal ancestry of Kavaratti islanders is closer to the northern and northwestern populations compared to the mainland Dravidian tribes on the southwest coast of India.

This result supports and is congruent with earlier findings by (Mustak et al., 2019) the clustering pattern observed in the maternal ancestry. Therefore, we can conclude that both paternal and maternal lineages share their ancestral roots in the North-west region. The fine structure of the paternal gene pool shows that although it is clustering broadly with the northern and northwestern groups in the multivariate Analysis (PCA).

Summary and Conclusion

The current study highlights the genetic ancestry of the Kavaratti Island population of Lakshadweep. Kavaratti island is one among the ten inhabited islands and it is large and the capital of Lakshadweep, union territory of India. It is located southwest of Kerala and 350 km away in the Arabian Sea. It is an isolated island and reside approximately 11,000 population and they practice endogamy. The genetic route of the Kavaratti Island population is not yet known. Therefore, this study tries to investigate the maternal and paternal ancestry, as well as the genetic affinity of Kavaratti Island with other islands of Lakshadweep, and with the neighbouring South Asian population.

In order to determine the maternal lineage of the Kavaratti Island population, the partial coding region sequences (14856- 16023) and hypervariable region (HVR-I and HVR-II) of mtDNA were examined. The maternal ancestry of Kavaratti Island has revealed two significant mtDNA ancestral lineages, Haplogroups R30b2 and M2a'b, which were commonly observed among the population and occupied the highest amount of the total mtDNA variation on Kavaratti Island. The abundance of macrohaplogroup M (51%) in this population resembles the mainland populations of the Indian subcontinent. Macrohaplogroup R accounted for 32% of the total distribution whereas U4 haplogroup occurred at 10% frequency. Viewed at the macro level, this population carries Indian-specific M and ancient N in relatively equal proportion. These two haplogroups contrast with the mainland tribal gene pool, which shows a significant difference between the two. Interestingly, such equal distribution is also a characteristic of Northern India.

The R30b2 founder lineage observed in the Kavaratti islanders, arising from the northwestern region of the Indian subcontinent during the Last Glacial Maximum period. The only other populations to carry this lineage are the Scheduled Castes from the Punjab, Jat Sikhs and Nairs, who are either found in the northwestern portion of the Indian subcontinent or have recent historical links with this region. In addition, we also find unique clusters of pre-LGM M2 and M33 clades sharing common ancestors with those found in the northwestern populations.

A similar trend is seen in the clustering pattern based on the M3 and M30 haplogroups that are prevalent in this region and date to the LGM period. We also report a rare occurrence of the U4 lineage dated to the LGM period that is otherwise only found in Europe. Its presence in Kavaratti islanders only suggests that the ancestors of the founders admixed with a diverse group before drastic demographic changes that occurred during and after the LGM. Our conclusions are based on the control region of the mitochondrial DNA and therefore must be further substantiated with whole mitogenomes.

Secondly, to investigate the origin of paternal lineages of Kavaratti populations, a total of 23 unique haplotypes were observed from the Y STR analysis. Our data show great diversity among the population of Kavaratti Island. The hg R was found to be the most prominent macro haplogroup of the Kavaratti population with a higher frequency of 23 %, followed by haplogroup L as the second most frequent haplogroup in the Kavaratti Island population, which characterizes (19%) of the total Y chromosome gene pool. Apart from the major two haplogroups observed in Kavaratti Island, there are a few more minor haplogroups observed, the minor haplogroup J accounts for (18%), Followed by O (12%) and E (7%). and the remaining haplogroups detected were less than (6%).

The present study has explored the maternal genetic affinity of Kavaratti Island with the other ten inhabited islands of Lakshadweep. Based on the haplotype frequencies, the Kavaratti Island population showed more varied (Fst = 0.858) when compared to the average pairwise differences (Fst) with other Lakshadweep islands. The stronger founder impact of R30 on these islands is mainly to blame for the enhanced diversity. Regarding Nei's genetic distance analysis for Kavaratti Island with the other islands, the genetic composition of the island turned out to be more similar to Amini, Minicoy and Kalpēni. It can be concluded that genetic admixture between Kavaratti and these islands occurred more frequently compared to other more distant islands. Interestingly, being a diverse group, Kavaratti and Amini retained the founder haplogroup R30 at ~30%, whilst Minicoy and Kalpēni lost this haplogroup gradually. Conversely, it can also be hypothesised that the northern islands including Amini, Agatti, Kiltān and Kavaratti were explored first by the R30 founders, and Minicoy was the latest.

To assess the relationship between founder haplogroups and island area, and to reconstruct the Paternal ancestral lineage of the Kavaratti islanders, we compared their gene pool with that of the other Lakshadweep populations. We used Principal Component Analysis (PCA) to visualise the clustering pattern among them. PCA shows a distinct cluster for Kavaratti. All the northern islands clustered together.

NJ Tree of Y-STR haplogroup showed Kavaratti clustered together with Minicoy and Androth. By these results, we can summarise that genetic admixture in between Kavaratti and these islands occurred more frequently than between Kavaratti and other islands. This resemblance may be due to the tight proximity of these islands to Kavaratti. The Kavaratti population is a more diverse group compared to Chetlat, Agatti and Kiltan. The result supports the hypothesis that the earliest settlers had a high frequency of Y-chromosome haplogroups R when they reached the Lakshadweep archipelago.

To determine the paternal genetic ancestry of the Kavaratti Population with the neighbouring South Asian population, Y-chromosome Hg frequencies were retrieved from the literature and distinguished based on geography and ethnicity. Comparing the maternal gene pools of the populations of the Kavaratti population with the mainland, study population were found to have an affinity with the Schedule Castes of Punjab and the Kodava, Koraga, Madia and Kathodi tribes. The Neighbour Joining (NJ) Tree plotted using Hamming's dissimilarity and the Andamanese root exhibits a similar clustering pattern. NMDS plot shows an interesting clustering pattern, the study observed that the Y-STR haplotypes of Kavaratti islanders show similarity with the North-western populations like Sindhi, Pathan and Gujjar did not observe any cluster with the south Indian populations from Tamil Nadu and Andhra Pradesh. Results suggest that different paternal lineage among the Kavaratti islanders indicates the paternal ancestry of Kavaratti islanders is closer to the northern and northwestern populations compared to the mainland Dravidian tribes on the southwest coast of India. Therefore, study conclude that both paternal and maternal lineages share their ancestral roots in the North-west region.

CONCLUSION

Kavaratti Island is the suitable genetic model to explore the genetic structure with relevance to their origin because of its isolated and endogamy practice in net shell, the following are the major findings of the present study,

1. Mitochondrial DNA analysis identified two major maternal founder haplogroups R30b2 and M2ab commonly observed in the north and northwestern parts of India with rare occurrence of U4 haplogroup lineage dated the LGM period.
2. Y-STR analysis identified two major paternal founder haplogroups R and L with minor haplogroups J and O. showing the paternal ancestry components of Kavaratti Island are more similar to closer to the northern and northwestern populations including Pakistan compared to the mainland Dravidian tribes on the southwest coast of India.

3. The Kavaratti population (Fst = 0.858) is a more diverse group compared to Chetlat (0.324), Bitra (0.564) and Kadmat (0.526). This enhanced diversity is primarily due to the stronger founder effect of R30 on these islands Genetic admixture between Kavaratti and a few neighbouring islands occurred compared to other more distant islands. Kavaratti and Amini retained the founder haplogroup R30 at ~30%, while Minicoy and Kalpeni lost this haplogroup gradually
4. Genetic admixture between Kavaratti and a few neighboring islands occurred more frequently compared to other more distant islands. Kavaratti and Amini retained the founder haplogroup R30 at ~30%, while Minicoy and Kalpeni lost this haplogroup gradually
5. From Bayesian analysis, we establish a time to the most recent common ancestor of 17,616 (95% HPD: 6,661-32,561) for R30b2 and the most recent common ancestor of ~12,000 years for R30a1c, whilst the R30a clade dates back to ~39,000 YBP.
6. The multivariate analysis, suggests the Kavaratti population belonged to a unique wave of migrants that neither belonged to the earliest Indian specific M2 carriers nor to the West Eurasian descendants. A peculiar feature in the Kavaratti cluster is the presence of M3, M30 and U2. M3 and M30 originated during or after the LGM event. This time depth is important, since the members of the Kavaratti cluster carry M3 and M30, leading us to the assumption that these populations had similar gene pools during the LGM recovery, after which bottleneck events took place.

OVERALL CONCLUSION

Both the maternal and paternal ancestry of Kavaratti Island show higher affinity with the populations of the Northern and North Western parts of India and Pakistan

References

1. Anthropological Survey of India (2021) *Genomic Diversity in People of India: Focus on mtDNA and Y-Chromosome polymorphism*. 1st edn. Springer Singapore. doi: 10.1007/978-981-16-0163-7.
2. Ballinger S. W., Schurr T. G., Torroni A., Gan Y.-Y., Hodge J. A., Hassan K., Chen K. H., Wallace D. C., (1992) "Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations". Genetics. 130. pp 139-152.
3. Bamshad, M., Kivisild, T., Watkins, W. S., Dixon, M. E., Ricker, C. E., Rao, B. B., ... & Papiha, S. S. (2001). Genetic evidence on the origins of Indian caste populations. Genome research, 11(6), 994-1004.
4. Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., ... & Majumder, P. P. (2003). Ethnic India: a genomic view, with special reference to peopling and structure. *Genome research*, *13*(10), 2277-2290.
5. Bittles A. H., (2002) "The impact of consanguinity on the Indian population". Indian Journal of Human Genetics. 8. pp 45.
6. Cann, R. L. (2001). Genetic clues to dispersal in human populations: retracing the past from the present. *Science*, *291*(5509), 1742-1748.
7. Chaubey, G., Metspalu, M., Kivisild, T., & Villems, R. (2007). Peopling of South Asia: investigating the caste–tribe continuum in India. *Bioessays*, *29*(1), 91-100.
8. Dür, A., Huber, N., & Parson, W. (2021). Fine-tuning phylogenetic alignment and haplogrouping of mtDNA sequences. *International Journal of Molecular Sciences*, *22*(11), 5747.
9. Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, *10*(3), 564-567.
10. Forbes, A. D. W. (1979) 'Sources towards a history of the Laccadive Islands', *South Asia J South Asian Stud.*, 2. doi: 10.1080/00856407908722989.
11. gounder Palanichamy, M., Sun, C., Agrawal, S., Bandelt, H. J., Kong, Q. P., Khan, F., ... & Zhang, Y. P. (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *The American Journal of Human Genetics*, *75*(6), 966-978.
12. Groeneveld, E. (2017). Early Human Migration. Ancient History Encyclopedia.
13. Hammer, Ø., Harper, D. A., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia electronica*, *4*(1), 9.

14. Huber, N., Parson, W., & Dür, A. (2018). Next generation database search algorithm for forensic mitogenome analyses. *Forensic Science International: Genetics*, *37*, 204-214.
15. Joseph, T. (2018). How ancient DNA may rewrite prehistory in India. BBC News, 30.
16. Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., ... & Villems, R. (1999). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current biology*, *9*(22), 1331-1334.
17. Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in ecology and evolution*, *6*(9), 1110-1116.
18. Majumder, P. P. (2001). Ethnic populations of India as seen from an evolutionary perspective. *Journal of biosciences*, *26*, 533-545.
19. Mellars, P., Gori, K. C., Carr, M., Soares, P. A., & Richards, M. B. (2013). Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. Proceedings of the National Academy of Sciences, 110(26), 10699-10704.
20. Metspalu, M., Kivisild, T., Metspalu, E., Parik, J., Hudjashov, G., Kaldma, K., ... & Villems, R. (2004). Most of the extant mtDNA boundaries in south and southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC genetics*, *5*(1), 1-25.
21. Misra V., (2001) "Prehistoric human colonization of India". Journal of Biosciences. 26. pp 491-531.
22. Mustak, M. S., Rai, N., Naveen, M. R., Prakash, S., Carlus, S. J., Pasupuleti, N., ... & Thangaraj, K. (2019). the peopling of Lakshadweep Archipelago. *Scientific Reports*, *9*(1), 1-7.
23. Okonechnikov, K., Golosova, O., Fursov, M., & Ugene Team. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, *28*(8), 1166-1167.
24. Ottolenghi C., Gadgil M., Cavalli-Sforza L. L., (1995) "Demographic history of India and mtDNA-sequence diversity". Am J Hum Genet. 56. pp 979992.
25. Palanichamy, M. G., Mitra, B., Zhang, C. L., Debnath, M., Li, G. M., Wang, H. W., ... & Zhang, Y. P. (2015). West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. *Human genetics*, *134*, 637-647.
26. Pasupuleti, N., & Mustak, M. S. (2020). Ancient DNA: revisiting past through population genomics. *Polymorphism*, *5*, 45-52.

27. Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., ... & McElreavey, K. (2004). Where west meets east: the complex mtDNA landscape of the southwest and Central Asian corridor. *The American Journal of Human Genetics*, *74*(5), 827-845.
28. Ranith, R., Senthilnathan, L., Machendiranathan, M., Thangaradjou, T., Sasamal, S. K., & Choudhury, S. B. (2017). Mapping sites of reef vulnerability along lagoons of Lakshadweep archipelago, Indian Ocean. *Environmental Monitoring and Assessment*, *189*(10), 494.
29. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular biology and evolution*, *34*(12), 3299-3302.
30. Sahoo, S., & Kashyap, V. K. (2006). Phylogeography of mitochondrial DNA and Y-Chromosome haplogroups reveal asymmetric gene flow in populations of Eastern India. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 131(1), 84-97.
31. Schlebusch, C. M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., ... & Lombard, M. (2017). Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science, 358(6363), 652-655.
32. Singh, G. *et al.* (2016) 'Genetic differentiation and population structure of five ethnic groups of Punjab (North-West India)', *Molecular Genetics and Genomics*, 291(6), pp. 2055–2063. doi: 10.1007/s00438-016-1239-3.
33. Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Molecular biology and evolution*, *38*(7), 3022-3027.
34. Thanseem, I., Thangaraj, K., Chaubey, G., Singh, V. K., Bhaskar, L. V., Reddy, B. M., ... & Singh, L. (2006). Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. *BMC genetics*, *7*, 1-11.
35. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H. J., ... & Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic acids research*, *44*(W1), W58-W63.
36. Thangaraj K, Chaubey G, Singh VK, et al (2006) In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup "M" in India. BMC Genomics 7:151. doi: 10.1186/1471-2164-7-151

Publications and Presentations

- **Related Publications**

1. **Alnoman Mundher Tayyeh**, Jaison J Sequeira, Lomous Kumar, Idrees Babu, George van Driem and Mohammed Shafiul Mustak (2023). The Maternal ancestry of the Kavaratti Islanders and the Last Glacial Maximum Aftermath. **Molecular Genetics and Genomics (MGAG).**

- **Unrelated Publications**

1. Fadam Abdoon, Mustafa J. Bichan, Abdelrahman Mohamed Ibrahim, **Alnoman Mundher Tayyeh**, Medikondu Kishore, and Azam Abdelhakeem Khalid. Ternary Complexation Process for New Spectrophotometric Assay of Levodopa using Ni(II) and 2,3-Diaminopyridine. Advances in Materials Science and Engineering 2022(3):1-8 mailto:https://www.hindawi.com/journals/amse/2022/4915162/

2. Saja Sumiea Anaz, **Alnoman Mundher Tayyeh**. Characteristics of trauma patients with multidrug-resistant bacteria from an epidemiological, clinical, and microbiological perspective. Tamjeed Journal of Healthcare Engineering and Science Technology (TJHEST), Volume 1 (2): 2023. mailto:https://www.tamjed.com/index.php/TJHEST/article/view/20

3. Abdulsattar Abdullah Hamad, Faris Maher Ahmed, Labesh Kumar, Sumalatha Donipati, Talluri Sreekrishna, Din Bandhu, Bathina Rajesh Kumar, **Alnoman Mundher Tayyeh.** Development of Cellulose nanocomposites for Electromagnetic shielding applications by Using Dynamic Network, Journal of Process Mechanical Engineering. mailto:https://journals.sagepub.com/doi/abs/10.1177/09544089231202913

Presentation at National/ International conferences

1. **Alnoman M T**, Jaison J S and Mustak M S, Presented a Poster intitle **(The peopling of the Kavaratti Island of Lakshadweep, India - A mitochondrial analysis)** at the 7th International conference of the collage of science for biological sciences (CBSc - 2022) in September 20-21, 2022. Mustansiriyah University, Baghdad - IRAQ.
2. **Alnoman M T**, Jaison J S and Mustak M S, Presented a Poster on **(The genetic analysis of mtDNA of Kavaratti Island population of Lakshadweep, India.)** at the International Conference on Frontiers in Nutrition, Medical Genomics, and Drug Discovery (Indix'22), Vignan University, on 31 Oct - 02 Nov 2022.
3. **Alnoman M T**, Jaison J S, Nagarjuna P, Thangaraj K, Mustak M S, presented a Poster on **(The genetic history and composite structure of three different tribes of Amini Island of Lakshadweep- A whole mitochondrial analysis)** at the Indian Anthropology Congress-2022, organized by the Department of Anthropology, University of Hyderabad in the collaboration with Indian National Confederation and Academy of Anthropologists, on 21- 23 February 2022.
4. **Alnoman Mundher Tayyeh**, Jaison J Sequeira and Mohammed S Mustak, presented a Poster on **(Phylogenetic analysis of mtDNA lineages in Kavaratti Island population Lakshadweep, India)** at the 9th Annual Conference of the Society for Mitochondrial Research & Medicine "Mitochondria in Biology and Medicine", Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India on 21 – 23 June 2023

Research Scholar	Research Guide
Alnoman Mundher Tayyeh	Dr. M. S. Mustak